# MANAV RACHNA INTERNATIONAL UNIVERSITY

## Research Plan

## For

## Ph. D. Programme 2009-10

## Robust Automatic Speaker Recognition System


**DEPARTMENT OF ELECTRONICS & COMMUNICATION**

**FACULTY OF ENGINEERING & TECHNOLOGY**

**Submitted by:**

**Name: Geeta Nijhawan**

**Registration No.: 09019990031**

**Supervisor:**

**Name: Dr M. K. Soni**

**Designation: Executive Director and Dean (FET)**

**Co-Supervisor:**

**Not Applicable**

## ABSTRACT

This research work aims at designing both text-dependent and text-independent speaker recognition system based on mel frequency cepstral coefficients (MFCCs) and voice activity detector (VAD). VAD has been employed to suppress the background noise and distinguish between silence and voice activity. MFCCs will be extracted from the detected voice sample and will be compared with the database for recognition of the speaker. A new criterion for detection is proposed which is expected to show very good performance in noisy environment. The system will be implemented on MATLAB platform and a new approach for designing a voice activity detector (VAD) has been proposed. The effectiveness of the proposed system can be proved by comparative analysis of the proposed design approach with the artificial neural networks technique. In past few years there has been lot of work that has proved artificial neural networks (ANN's) as a powerful tool for speaker recognition. The performance of both the systems will be evaluated under different noisy environments and in different languages and emotions. The overall efficiency of the proposed speaker recognition system depends mainly on the detection criteria used for recognizing a particular speaker. Global optimization techniques like Genetic Algorithm (GA), Particle Swarm Optimization (PSO) etc. can prove very useful in this context and hence for setting up of the detection criteria Genetic Algorithm will be employed.



**Keywords**: Speaker recognition, acoustic processing, feature extraction, MFCC, voice activity detector, feature matching, Euclidean distance, neural network, optimization techniques.

# CONTENTS


| S.No. | Description | Page No. |
|---|---|---|
| 1. | Introduction | 1 |
| 2. | Literature Review | 2 |
| 3. | Description of broad area | 4 |
| 4. | Objectives of the study | 8 |
| 5. | Methodology | 8 |
| 6. | Proposed output of the research | 10 |
| 7. | References | 12 |

# INTRODUCTION

Development of speaker recognition system began in early 1960's with the exploration into voiceprint analysis. The detection efficiency of speaker recognition systems gets severely affected in the presence of noise. This fact ensured to derive a more reliable method. Speaker recognition is the process of recognizing the speaker from the database based on some characteristics in the speech wave. Most of the speaker recognition systems contain two phases. In the first phase feature extraction is done. The unique features from the voice data are extracted which are used latter for identifying the speaker. The second phase is feature matching in which we compare the extracted voice data features with the database of known speakers. The overall efficiency of the system depends on how efficiently the features of the voice are extracted and the procedures used to compare the real time voice sample features with the database.

For security application to crime investigations, speaker recognition is one of the best biometric recognition technologies. We can give our speech signal as password to the lock system of our home, locker, computer etc. Speaker recognition can also be helpful in verifying voice of criminal from the audio tape of telephonic conversations. The main advantage of biometric password is that there is nothing like forgetting or misplacing it.

Voice biometric compared to other biometric is user friendly, cost-effective, convenient and secure. It finds application in the recognition of telephone numbers, personal identification numbers and credit card numbers.

The modern speaker recognition systems are designed keeping in mind that it should have high accuracy, low complexity and easy calculation. Hidden Markov Model (HMM) technique has proved effective for both the isolated word and continuous speech recognition; however it does not address discrimination and robustness issues for classification problems. The acoustic analysis based on MFCC which represents the ear model [1], has given good results in speaker recognition. Background noise and microphone used also effect the overall performance of the system [2].

Speaker recognition systems contain three main modules:

(1) Acoustic processing

(2) Features extraction or spectral analysis

(3) Recognition.

All the three modules are shown in **Fig. 1** and are explained in detail in the subsequent sections.

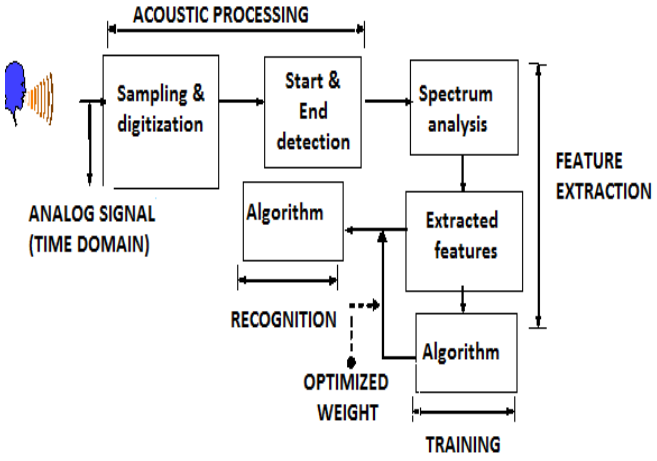


**Fig.1 Basic structure of speaker recognition system**

For more than four decades, efforts have been made to make speaker recognition methods more efficient and it is still an active area for research and development. Many approaches like human aural and spectrogram comparisons, simple template matching, dynamic time-warping approaches, and modern statistical pattern recognition approaches, such as neural networks and Hidden Markov Models (HMMs) have been used. Many techniques have been used for speaker recognition including Vector Quantisation(VQ), Gaussian Mixture Modeling (GMM),neural networks and genetic algorithms [3].

## LITERATURE REVIEW

Research has been focused on Feature based Recognition Systems. Using features from speech based sources it has been tried to create a reliable, robust and efficient recognition system. However, the complexity of such a system increases because of variations caused due to differences in individual speaker characteristics, emotion variations and noise disturbances.

Text-dependent methods use template-matching techniques. Feature vectors are extracted from the input speech. Dynamic time warping (DTW) algorithm is used to align the time axes of the input speech and each reference template or model of the registered speakers [4]. From the beginning to the end of the speech, the degree of similarity between them is calculated. Statistical variation in spectral features can be modeled by Hidden Markov Model (HMM).

HMM-based methods are extensions of the DTW-based methods .A new technique for computing verification scores using multiple verification features from the list of scores for a target speaker's speech was introduced by Park, A (2001)[5].This technique was compared to the baseline logarithmic

likelihood ratio verification score using global GMM speaker models .It gave no improvement in verification performance.

Zhou, L (2000) used neural networks and fuzzy techniques [8]. A recognition rate of 92.2% was achieved for a speaker independent speech recognition system. The tests were conducted for a large collection of speech templates of Chinese digits 0—9 which was taken from the persons from different areas and in noisy environment.

Moonasar, V, Venayagamoorthy, G (2002) proposed a speaker verification system with the use of a committee of neural networks rather than the conventional single network decision system. Supervised Learning Vector Quantization (LVQ) were used as recognizer .There was reduction in recognition rate with increase in number of speakers to be recognized.Hybrid feature parameter vectors were made using Linear Predictive Coding (LPC) and Cepstral signal processing techniques .

The most commonly used acoustic vectors are Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC) and Perceptual Linear Prediction Cepstral (PLPC) coefficients and zero crossing coefficients (Yegnanarayana *et al*, 2005; Vogt *et al*, 2005). The spectral information is obtained from a short time windowed segment of speech.These feature vectors differ mainly in the power spectrum representation. A modification of Mel-Frequency Cepstral Coefficient (MFCC) feature has been proposed (Saha and Yadhunandan, 2000.Multi-dimensional F-ratio is used as performance measure to compare discriminative ability. Bark scale also gives same performance in speech recognition experiments as MFCC (Aronowitz *et al*, 2005) .They are effective for text dependent speaker verification systems. Kumar *et al*, (2010), Ming *et al*, (2007) proposed Revised Perceptual Linear Prediction Coefficients (RPLP), in which coefficients were obtained from combination of MFCC and PLP. These coefficients are useful for identifying the spoken language.

Earlier work on speaker recognition used direct template matching between training and testing data. Similarity measure between training and testing feature vectors was used in the direct template matching.Techniques like spectral or Euclidean distance or Mahalanobis distance is used (Liu *et al*, 2006).But as the number of feature vectors increases the method becomes time consuming. To decrease the number of training feature vectors we use clustering. The cluster centres form code vectors and the set of code vectors is called codebook. K-means algorithm is the commonly used codebook generation algorithm (Mporas *et al*, 2007; Ming *et al*, 2007). In 1985,Soong *et al*. used the VQ-LBG algorithm.The performance of speaker recognition systems in neural network based networks were also examined (Clarkson *et al*., 2006). Continuous probability measures are created using Gaussian mixtures models (GMMs) (Krause and Gazit, 2006). In 1995, Reynolds proposed Gaussian mixture modeling (GMM) classifier for speaker recognition task (Krause and Gazit, 2006; Clarkson *et al*, 2006).This is the widely used probabilistic technique in speaker recognition. The GMM needs sufficient data to model the speaker (Aronowitz *et al*, 2005). In the GMM modeling

technique, the distribution of feature vectors is modeled by the mean, covariance and weight.The performance of GMM is much better than other techniques.

Various researchers are still trying to improve the peformance of speaker recognition systems so as to achieve better peformance .Use of various existing optimization techniques namely genetic algorithm, particle swarm optimization, neural networks etc can come handy in improving the performance .

## DESCRIPTION OF BROAD AREA/TOPIC

Speaker recognition is the process of recognizing the speaker from the database based on some characteristics in the speech wave. Most of the speaker recognition systems contain two phases. In the first phase feature extraction is done. The unique features from the voice data are extracted which are used latter for identifying the speaker. The second phase is feature matching in which we compare the extracted voice data features with the database of known speakers[9]. Each module will be discussed in detail in later sections.

### 1. ACOUSTIC PROCESSING

Acoustic processing is sequence of processes that receives analog signal from a speaker and convert it into digital signal for digital processing. Human speech frequency usually lies in between 300Hz-8000kHz [10].Therefore 16kHz sampling size can be chosen for recording which is twice the frequency of the original signal and follows the Nyquist rule of sampling [11].The start and end detection of isolated signal is a straight forward process which detect abrupt changes in the signal through a given threshold energy. The result of acoustic processing would be discrete time voice signal which contains meaningful information. The signal is then fed into spectral analyser for feature extraction.

### 2. FEATURE EXTRACTION

Feature Extraction module provides the acoustic feature vectors used to characterize the spectral properties of the time varying speech signal such that its output eases the work of recognition stage. A small amount of speaker specific information in the form of feature vectors from the input voice signal is extracted and it is used as a reference model representing each speaker's identity. A general block diagram of speaker recognition system is shown in **Fig 2 [12]**.

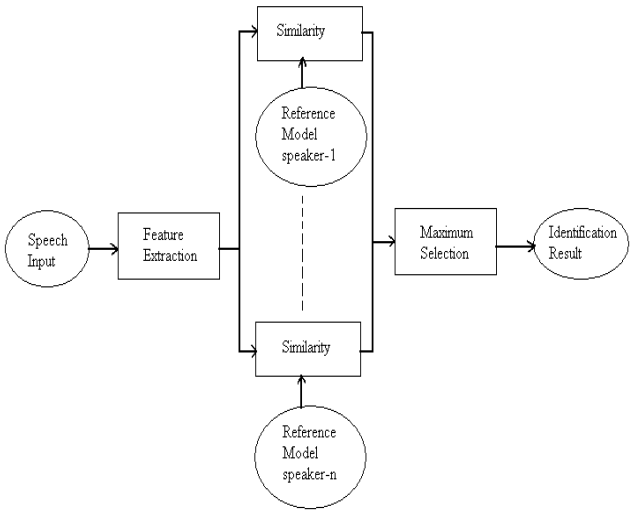


**Fig.2 Speaker recognition system**

It is clear from the above diagram that the speaker recognition is a 1:N match where one unknown speaker's extracted features are matched to all the templates in the reference model for finding the closest match. The speaker feature with maximum similarity is selected.

### *A. MFCC Extraction*

Mel frequency cepstral coefficients (MFCC) is probably the best known and most widely used for both speech and speaker recognition. A mel is a unit of measure based on human ear's perceived frequency. The mel scale is approximately linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz[13]. The approximation of mel from frequency can be expressed as-

$$mel(f) = 2595*\log(1+f/700) \qquad \text{--------(1)}$$

where f denotes the real frequency and mel(f) denotes the perceived frequency. The block diagram showing the computation of MFCC is shown in **Fig. 3**.

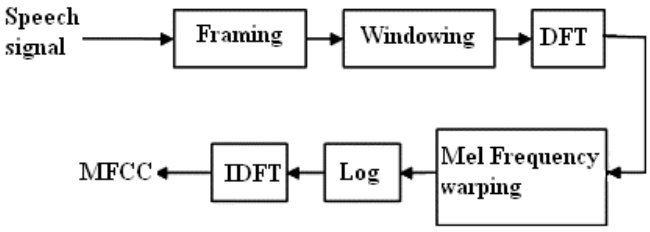


**Fig.3 MFCC Extraction**

In the first stage speech signal is divided into frames with the length of 20 to 40 ms and an overlap of 50% to 75%. In the second stage windowing of each frame with some window function is done to minimize the discontinuities of the signal by tapering the beginning and end of each frame to zero. In time domain window is

point wise multiplication of the framed signal and the window function. A good window function has a narrow main lobe and low side lobe levels in their transfer function. In our work hamming window is used to perform windowing function [14]. In third stage DFT block converts each frame from time domain to frequency domain. In the next stage mel frequency warping is done to transfer the real frequency scale to human perceived frequency scale called the mel-frequency scale. The new scale spaces linearly below 1000Hz and logarithmically above 1000Hz. The mel frequency warping is normally realized by triangular filter banks with the center frequency of the filter normally evenly spaced on the frequency axis. The warped axis is implemented according to equation 1 so as to mimic the human ears perception. The o/p of the ith filter is given by-

$$y(i) = \sum_{j=1}^{N} s(j)\Omega_i(j) \qquad \text{---------- (2)}$$

S(j) is the N-point magnitude spectrum (j =1:N) and Ωi(j) is the sampled magnitude response of an M-channel filter bank (i =1:M). In the fifth stage Log of the filter bank output is computed and finally DCT (Discrete Cosine Transform) is computed. The MFCC may be calculated using the equation-

$$C_s(n,m) = \sum_{i=1}^{M} (\log Y(i)) \cos[i \frac{2\pi}{N'} n] \qquad \text{--------- (3)}$$

where N' is the number of points used to compute standard DFT.

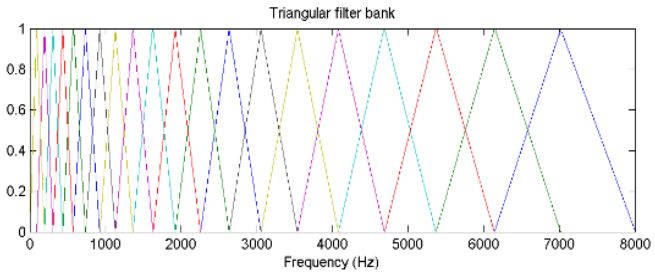


**Fig.4 Triangular filter bank**

### *B. Voice Activity Detector*

Voice Activity Detector (VAD) has been used to primarily distinguish speech signal from silence[15]. VAD compares the extracted features from the input speech signal with some predefined threshold. Voice activity exists if the measured feature values exceed the threshold limit, otherwise silence is assumed to be present. Block diagram of the basic voice activity detector used in this work is shown in **Fig. 5.**

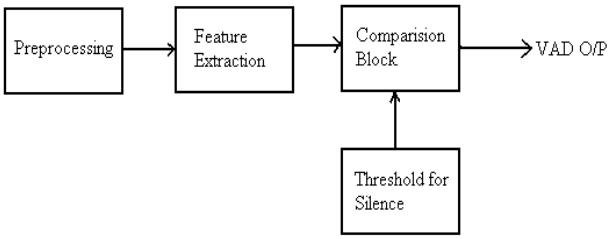


**Fig. 5 VAD block diagram**

The performance of the VAD depends heavily on the preset values of the threshold for detection of voice activity. The VAD proposed here works well when the energy of the speech signal is higher than the background noise and the background noise is relatively stationary. The amplitude of the speech signal samples are compared with the threshold value which is being decided by analyzing the performance of the system under different noisy environments.

**3. FEATURE MATCHING**

### *A. Using Euclidean Distance*

A sequence of feature vectors $\{x_1, x_2,\ldots,x_n\}$for unknown speakers are extracted. These are compared with the feature vectors already stored in the database. For each pair of feature vectors a distortion measure is calculated. The speaker with the lowest distortion is chosen[16],[17].

Thus, each feature vector of the input is compared with all the codebooks. The codebook with the least average distance is chosen to be the best. The formula used to calculate the Euclidean distance can be defined as follows:

Let us take two points P = ($p_1$, $p_2 \ldots p_n$) and Q= ($q_1$, $q_2$...qn). The Euclidean distance between them is given by

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad \text{-------- (4)}$$

The speaker with the lowest distortion distance is chosen as the unknown person.

### *B. Neural Networks (NN)*

Several popular pattern matching techniques: HMM, GMM, DTW, VQ, NN are being used for Speaker Recognition. We have chosen neural network as recognizer.

In the recognition phase, the neural networks are trained to learn the mapping from the features extracted from the pre-separated speech to those extracted from the close-talking microphone speech

signal. The outputs of the neural networks are then used to generate acoustic features, which are subsequently used in acoustic model adaptation and system evaluation [18].

## OBJECTIVES OF THE STUDY

Automatic speaker recognition works on the principle that a person's speech exhibits characteristics that are unique to the speaker. Speech signals in training and testing sessions cannot be same due to many facts such as people's voice change with time, health conditions, speaking rates, etc. Acoustical noise and variations in recording environments present a challenge to speech recognition [19].The challenge would be to make the system "**Robust**". If the recognition accuracy does not degrade significantly, the system is called "**Robust**".

.

**The objectives of this research work are:**

1. Develop a new text-dependent and text-independent speaker recognition framework with the help of MFCC and VAD.

2. Dynamically train the speaker recognition system with clean and noisy (additive and convolutive) speech signals. Each time a new speech signal is input to the system, additive white Gaussian noise at different values of SNR and echo with varying values of delay are added to the clean speech signals.

3. Investigate the performance of the proposed text-independent and text-dependent speaker recognition systems under noisy environments.

4. Compute the accuracy rates of identifying the test speaker in clean and noisy environments using the designed speaker recognition model and compare it with the artificial neural network based speaker recognition technique.

5. To analyze the best method of removing background noise in voice signal.

## METHODOLOGY

Most of the speaker recognition systems contain two phases. First phase is feature extraction in which the unique features from the voice data are extracted which are used latter for identifying the speaker. In the second phase is feature matching and this phase comprises of the actual procedures carried out for identifying the

speaker by comparing the extracted voice data features with the database of known speakers. The overall efficiency of the system depends on the fact that how efficiently the features of the voice are extracted and the procedures used for comparing the real time voice sample features with the database [20].

The following steps will be performed:

a) Voice will be recorded using microphone

b) Voice activity detection to be performed on the extracted voice

c) Feature extraction using MFCC

d) Speaker recognition using Euclidean distance

e) Compare the result obtained in (d) using Neural Network

f) Calculate % error for (d) and (e)

g) Display on serial port

**Data:** This work focuses on developing a system that uses the speech signal as a recognition system. The speech signal will be recorded using microphone. The signal is text dependent, where speakers will utter the words which will form a database. Different speaker will generate different speech waves.

**Tools:** The main tool that will be used in this research is MATLAB software. The MATLAB DSP (Digital Signal processing) toolbox and neural network toolboxes will be used to develop the programs in the software. A GUI will be designed in MATLAB for speaker recognition.

**Hardware:** The hardware that will be used in this research is:

1. Laptop
2. Intel Pentium Core 2 Duo 1.6GHz
3. USB PC Microphone

**Fig.6** shows the flow chart of Automatic Speaker Recognition System.

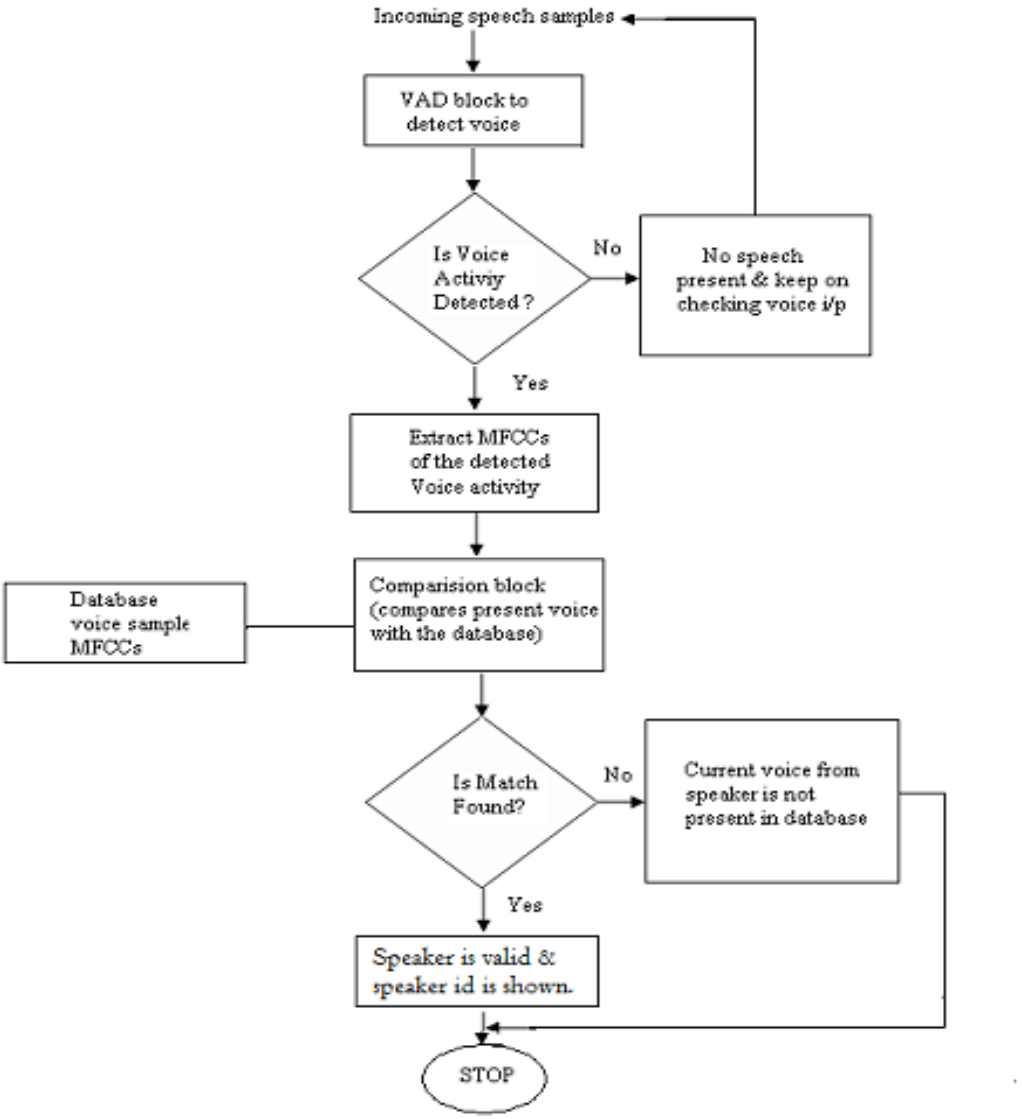
Incoming speech samples
VAD block to detect voice
Is Voice Activiy Detected ?
No
No speech present & keep on checking voice i/p
Yes
Extract MFCCs of the detected Voice activity
Database voice sample MFCCs
Comparision block (compares present voice with the database)
Is Match Found?
No
Current voice from speaker is not present in database
Yes
Speaker is valid & speaker id is shown.
STOP

**Fig 6: Flow chart of Speaker Recognition System**

## PROPOSED OUTPUT OF THE RESEARCH

The complete system will consist of software coded in matlab with graphical user interface, a mic for capturing voice based data and a hardware circuit connected to the computer via serial port used for operating a lock and delivering the result on LCD.

As soon as the system is activated, the microphone connected to a computer will start capturing voice based signals and converting them to electrical signals that can be saved and analyzed.

Coded in MATLAB the system will analyze the data captured by microphone for white noise and for background data that will be differentiated by voice if it is below a specified threshold limit.

This data will be utilized to filter out the needed speech command from the complete voice signal having noise and background sound. The task will be accomplished by generating voice signals similar to noise and background sound but will be 180 degrees out of phase with them, so as that can be cancelled resulting in only the needed speech command.

Once the voice command is successfully extracted from the complete signal, this will be then analyzed, extracting various parameters needed for successful comparison to the database speech.

The extracted features will be:

1. Base frequencies present in the signal
2. The amplitude variation of the peaks
3. The energy envelope present in the signal

The above mentioned parameters will be compared with the parameters of the speech stored in database in the form of wave file. A threshold will be defined for each feature, if the comparisons made for each feature is under specified thresholds, then the result will be declared true otherwise false. In either case a data packet associated with the result will be sent over serial port (UART protocol), to the microcontroller.

The hardware part will consist of a microcontroller, Relay and 16x2 LCD. On receiving the message from the computer via serial port (UART protocol) this microcontroller will operate a relay and will flash a message on the LCD reporting the result either matched or unmatched. The relay output further can be used to operate a actuator to open or close a door.

**REFERENCES**


[1]Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal," Bangla Speech Recognition System using LPC and ANN",Seventh International Conference on Advances in Pattern Recognition,2009

[2]Amruta Anantrao Malode , Shashikant Sahare,2012 Advanced Speaker Recognition, International Journal of Advances in Engineering & Technology ,Vol. 4, Issue 1, pp. 443-455.

[3] A.Srinivasan,"Speaker Identification and verification using Vector Quantization and Mel frequency Cepstral Coefficients", Research Journal of Applied Sciences, Engineering and Technology4(I):33-40,2012.

[4]B. Peskin, J. Navratil, J. Abramson, D. Jones, D. Klusacek, D.A. Reynolds, and B. Xiang, "Using prosodic and conversational features for high-performance speaker recognition," in Int. Conf. Acoust., Speech, Signal Process., vol. IV, Hong Kong, Apr. 2003, pp. 784-7.

[5] B. Yegnanarayana, S.R.M. Prasanna, J.M. Zachariah, and C.S. Gupta, "Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system," *IEEE Trans. Speech Audio Process.* , vol. 13(4), pp. 575-82, July 2005.

[6] B. Yegnanarayana, K. Sharat Reddy, and S.P. Kishore, "Source and system features for speaker recognition using AANN models," in proc. Int. Conf. Acoust., Speech, Signal Process., Utah, USA, Apr. 2001.

[7] Ch.Srinivasa Kumar, Dr. P. Mallikarjuna Rao, 2011,Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm, International Journal on Computer Science and Engineering,Vol. 3 No. 8 ,pp:2942-2954

[8] C.S. Gupta, "Significance of source features for speaker recognition," Master's thesis, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India, 2003.

[9] D.A. Reynolds, "Experimental evaluation of features for robust speaker identification," IEEE Trans. Speech Audio Process., vol. 2(4), pp. 639-43, Oct. 1994.

[10] Fu Zhonghua; Zhao Rongchun; "An overview of modeling technology of speaker recognition", IEEE Proceedings of the International Conference on Neural Networks and Signal Processing Volume 2, Page(s):887 – 891, Dec. 2003.

[11] F.K. Soong, A.E. Rosenberg, L.R. Rabiner, and B.H. Juang, "A Vector quantization approach to speaker recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* , vol. 10, Detroit, Michingon, Apr. 1985, pp. 387-90.

[12] Gabriel Zigelboim and Dr Ilan D. Shallom," A comparison Study of Cepstral Analysis with Applications to Speech Recognition", International Conference on Information Technology: Research and Education,2006

[13] Geeta Nijhawan, M.K. Soni," A Comparative Study of Two Different Neural Models For Speaker Recognition Systems", International Journal of Innovative Technology and Exploring Engineering,ISSN: 2278-3075, Volume-1,Issue-1,June 2012

[14] Harry Wechsler, Vishal Kakkad, Jeffrey Huang, Srinivas Gutta, V. Chen, "Automatic Video-based Person Authentication Using the RBF Network" First International Conference on Audio- and Video-Based Biometric Person Authentication, 1997 pages 85-92.

[15] Hui Kong, Xuchun Li, Lei Wang, Earn Khwang Teoh, Jian-Gang Wang, Venkateswarlu, R "Generalized 2D principal component analysis", Proc. 2005 IEEE International Joint Conference on Volume 1, Aug. 2005.

[16] John G. Proakis and Dimitris G. Manolakis, "Digital Signal Processing", New Delhi: Prentice Hall of India. 2002.

[17] Jayanna HS, Mahadeva Prasanna SR. Analysis, Feature Extraction, Modeling and Testing Techniques for Speaker Recognition. IETE Tech Rev, Year 2009, Vol 26, Issue 3, Pg181-90

[18] Khalifa, O.O, et al, "Speech coding for Bluetooth with CVSD algorithm", Proc. RF and Microwave Conference. Selangor, Malaysia, Page(s):227 – 229, 5-6 Oct. 2004

[19] L. Rabiner, and B.H. Juang, Fundamentals of Speech Recognition. Singapore:Pearson Education, 1993.

[20] Md Sah Bin Hj Salam, Dzulkifli Mohamad Sheikh Hussain Shaikh Salleh," Temporal Speech Normalization Methods Comparison in Speech Recognition Using Neural Network", International Conference of Soft Computing and Pattern Recognition, 2009

[21] Md. Rashidul Hasan,Mustafa jamil,Md. Golam Rabbani Md Saifur Rahman,Speaker Identification Using Mel Frequency Cepstral coefficients,3rd international Conference on Electrical & Computer Engineering,ICECE 2004,28-30 December 2004,Dhaka ,Bangladesh

[22] M.J. Carey, E.S. Parris, H. Lloyd-Thomas, and S. Bennett, "Robust prosodic features for speaker identification," in proc. Int. Spoken Language Process., Philadelphia, PA, USA, Oct. 1996.

[23] M.K. Sonmez, E. Shriberg, L. Heck, and M. Weintraub, "Modeling dynamic prosodic variation for speaker verification," in proc. Int. Spoken Language Process., Sydney, NSW, Australia, Nov-Dec. 1998.

[24] Parson, T.W, "Voice and Speech Processing", New York, United States of America: McGraw-Hill. 294, 1987.

[25] P. Thevenaz, and H. Hugli, "Usefulness of the LPC-residue in text- independent speaker verification," Speech Communication, vol. 17, pp. 145-57, 1995

[26] Premakanthan, P.; Mikhael, W.B., "Speaker verification/recognition and the importance of selective feature extraction: review", Proceedings of the 44th IEEE 2001 Midwest Symposium on Circuits and Systems, 2001. MWSCAS 2001. Volume 1, 14-17 Page(s):57 –61. Aug. 2001

[27] Rudra Pratap. Getting Started with MATLAB 7. New Delhi: Oxford University Press, 2006

[28] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-34, pp. 52-9, Feb. 1986.

[29] Sasaoki Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Trans. Acoust., Speech, Signal Process., vol. 29(2), pp. 254-72, Apr. 1981.

[30] Seddik, H.; Rahmouni, A.; Sayadi, M.; "Text independent speaker recognition using the Mel frequency cepstral coefficients and a neural network classifier"First International Symposium on Control, Communications and Signal Processing, Proceedings of IEEE 2004 Page(s):631 – 634.

[31] S.R.M. Prasanna, C.S. Gupta, and B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech", Speech Communication, vol. 48, pp. 1243-61, 2006.

[32] Sumithra, M. G. "A New Speaker Recognition System with Combined Feature Extraction Techniques", Journal of Computer Science

[33] Vibha Tiwari,"MFCC and its applications in speaker recognition",International Journal on Emerging Technologies1(I):19-22(2010)

[34] Yongjin Wang and Ling Guan, “An investigation of speech-based human emotion recognition”, IEEE 6th Workshop on Multimedia Signal Processing, 2004

[35] Young, S., “A review of large vocabulary continuous speech”, IEEE Signal Processing Magazine, v. 13, n 5, pp 45-57, 1996